Multidimensional Visual Analyser

Ožbej Golob

Multidimensional Visual Analyser

Ožbej Golob

Master's Thesis

to achieve the university degree of

Diplom-Ingenieur

Master's Degree Programme: Computer Science

submitted to

Graz University of Technology

Supervisor

Ao.Univ.-Prof. Dr. Keith Andrews
Institute of Interactive Systems and Data Science (ISDS)

Graz, 10 Nov 2021

Das Schreiben einer Masterarbeit: Richtlinien zur Erstellung einer Masterarbeit in Informatik

Keith Andrews B.Sc.

Masterarbeit

für den akademischen Grad

Diplom-Ingenieur

Masterstudium: Informatik

an der

Technischen Universität Graz

Begutachter

Ao.Univ.-Prof. Dr. Keith Andrews
Institute of Interactive Systems and Data Science (ISDS)

Graz, 10 Nov 2021

Diese Arbeit ist in englischer Sprache verfasst.

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The document uploaded to TUGRAZonline is identical to the present thesis.

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Das in TUGRAZonline hochgeladene Dokument ist mit der vorliegenden Arbeit identisch.

Date/Datum

Signature/Unterschrift

Abstract

Writing a thesis is a vast, overwhelming endeavour. There are many obstacles and false dawns along the way. This thesis takes a fresh look at the process and addresses new ways of accomplishing this daunting goal.

The abstract should concisely describe what the thesis is about and what its contributions to the field are (what is new). Market your contributions to your readership. Also make sure you mention all relevant keywords in the abstract, since many readers read *only* the abstract and many search engines index *only* title and abstract.

This thesis explores the issues concerning the clear structuring and the academic criteria for a thesis and presents numerous novel insights. Special attention is paid to the use of clear and simple English for an international audience, and advice is given as to the use of technical aids to thesis production. Two appendices provide specific local guidance.

Kurzfassung

Eine Masterarbeit zu schreiben ist ein ausgedehntes und schwieriges Unterfangen. Viele Hindernisse und falsche Ansätze säumen den Weg. Diese Arbeit stellt einen neuartigen Zugang zu diesem Vorgang dar, und zeigt neue Wege auf, dieses mühselige Ziel zu erreichen.

Die Kurzfassung sollte den Inhalt der Masterarbeit kurz und klar beschreiben und die eigenen Leistungen hervorheben (was ist neu). Der eigene Beitrag zum Themenbereich soll möglichst interessant dargestellt werden. Alle relevanten Begriffe sollten in der Kurzfassung vorkommen, da viele Interessenten *nur* die Kurzfassung lesen und viele Suchmaschinen *nur* die Kurzfassung indizieren.

Im Besonderen werden Punkte, welche die klare Strukturierung und die akademischen Kriterien für eine Arbeit betreffen, untersucht, wobei viele neue Erkenntnisse präsentiert werden. Für ein internationales Publikum soll dem Gebrauch von klarem und einfachem Englisch besondere Aufmerksamkeit gewidmet werden, weiters werden Hinweise auf technische Hilfsmittel zur Erstellung einer schriftlichen Arbeit angegeben. Zwei Anhänge decken spezifische Richtlinien ab.

Contents

List of Figures

List of Tables

List of Listings

Acknowledgements

I am indebted to my colleagues at the ISDS and the Know-Center who have provided invaluable help and feedback during the course of my work. I especially wish to thank my advisor, Keith Andrews, for his immediate attention to my questions and endless hours of toil in correcting draft versions of this thesis.

Special mention goes to Christian Gütl, Irene Isser, and Josef Moser for their help in translating the thesis abstract into German and to Bernhard Zwantschko and Didi Freismuth for ample supplies of coffee. Please remember to replace this tongue-in-cheek acknowledgements section with your own version!

Last but not least, without the support and understanding of my wife, pleasant hours with my girlfriend, and the tolerance of my friends, this thesis would not have been possible.

Keith Andrews

Graz, Austria, 10 Nov 2021

Credits

I would like to thank the following individuals and organisations for permission to use their material:

- The thesis was written using Keith Andrews' skeleton thesis [Andrews 2021].
- Figure is used with kind permission of Keith Andrews, Graz University of Technology.

Chapter 1

Introduction

Intro

Chapter 2

Multidimensional Visual Analysis

Multidimensional visual analysis (MVA) is a powerful tool for exploring and understanding complex datasets. By using visual representations of data, such as graphs, charts, and maps, analysts can quickly and easily identify trends, patterns, and relationships within the data. This can provide valuable insights and help decision-makers make informed choices based on the information at hand.

In recent years, the availability of large and complex datasets has increased dramatically, making it increasingly difficult for analysts to make sense of the data using traditional methods. MVA offers a solution to this problem by allowing analysts to quickly and easily explore the data and identify key trends and patterns.

Through the use of advanced visualization techniques, analysts can create interactive and engaging visualizations that help to bring the data to life. These visualizations can be easily shared with others, allowing for collaboration and discussion around the data.

2.1 Multidimensional Visual Analysis Approaches

MVA approaches are methods and techniques used to analyze and understand complex data sets using visual representations. These approaches typically involve the use of specialized software or tools that allow analysts to create and manipulate graphical representations of the data in order to uncover patterns, trends, and relationships. In this section, we will review some popular MVA approaches.

2.1.1 Scatter Plots

A scatter plot is a type of graph that is used to display the relationship between two numerical variables, to identify any potential trends or patterns in the data, and to identify outliers. It uses dots or markers to represent the values of the two variables, and position of each dot on the graph indicates the value of the two variables in a single observation.

To create a scatter plot, the values of one variable are plotted on the x-axis (horizontal axis) and the values of the other variable are plotted on the y-axis (vertical axis). The resulting graph will show a set of dots, with each dot representing a single observation. If there is a positive relationship between the two variables, the dots will tend to form a diagonal line that slopes upwards from left to right. If there is a negative relationship, the dots will tend to form a diagonal line that slopes downwards from left to right. If there is no relationship between the two variables, the dots will be scattered randomly across the graph.

2.1.2 High Dimensional Projections

Visualizing high-dimensional data can be challenging because it is difficult for the human brain to comprehend more than three dimensions. High dimensional projections are techniques that are used to reduce the number of dimensions in the data and represent it in a way that is easier to understand and interpret. High dimensional projections can be further split into *linear* and *non-linear* projections.

2.1.2.1 Principal Component Analysis

Principal Component Analysis (PCA) is a linear projection. PCA uses linear algebra to identify the underlying dimensions or factors in the data and project the data onto a lower-dimensional space [Abdi and Williams 2010].

2.1.2.2 Multi-Dimensional Scaling

Multi-Dimensional Scaling (MDS) is a non-linear projection. MDS uses a distance metric to preserve the distances between data points in the high-dimensional space and project the data onto a lower-dimensional space [Morrison et al. 2003].

2.1.2.3 t-Distributed Stochastic Neighbor Embedding

t-distributed stochastic neighbor embedding (t-SNE) is a non-linear projection. t-SNE uses a probabilistic model to preserve the local structure of the data in the high-dimensional space and project the data onto a lower-dimensional space [Van der Maaten and Hinton 2008].

2.1.3 Parallel Coordinates

Parallel coordinates are a type of chart that is used to visualize multi-dimensional data. In this type of chart, each data point is represented by a vertical line that extends across multiple axes. The axes are typically arranged in parallel. Each axis represents a different variable, and the position of a point on that axis indicates the value of that variable for the data point. This allows multiple variables to be compared and analyzed simultaneously [Inselberg and Dimsdale 1990].

2.1.4 Cluster Analysis

Cluster analysis is a type of statistical technique used to identify groups of similar objects within a data set. It is a commonly used method for exploratory data analysis, and is often used as a way to gain insight into the underlying structure of the data. In cluster analysis, the data is divided into groups, or clusters, based on the similarity of the objects within each group. This allows analysts to identify common patterns and trends within the data, and to gain a better understanding of the relationships between the objects in the data set [Duran and Odell 2013].

2.2 Multidimensional Visual Analysis Software

MVA software is a type of computer program that is designed to help analysts visualize and analyze complex data sets. These programs typically include a wide range of tools and features that allow users to create and manipulate graphical representations of the data, such as scatter plots, parallel coordinates, and heat maps. By using these tools, analysts can quickly and easily explore the data and gain insights that might not be immediately apparent from looking at the raw data. MVA software is commonly used in fields such as business, finance, and marketing to help make data-driven decisions and uncover hidden trends in the data. In this section, we will review some popular MVA software programs.

Bibliography

Abdi, Hervé and Lynne J Williams [2010]. *Principal component analysis*. Wiley interdisciplinary reviews: computational statistics 2.4 (2010), pages 433–459 (cited on page 4).

Andrews, Keith [2021]. *Writing a Thesis: Guidelines for Writing a Master's Thesis in Computer Science*. Graz University of Technology, Austria. 10 Nov 2021. https://ftp.isds.tugraz.at/pub/keith/thesis/ (cited on page xi).

Duran, Benjamin S and Patrick L Odell [2013]. *Cluster analysis: a survey*. Volume 100. Springer Science & Business Media, 2013 (cited on page 4).

Inselberg, Alfred and Bernard Dimsdale [1990]. *Parallel coordinates: a tool for visualizing multi-dimensional geometry*. Proceedings of the First IEEE Conference on Visualization: Visualization90. IEEE. 1990, pages 361–378 (cited on page 4).

Morrison, Alistair, Greg Ross, and Matthew Chalmers [2003]. *Fast multidimensional scaling through sampling, springs and interpolation*. Information Visualization 2.1 (2003), pages 68–77 (cited on page 4).

Van der Maaten, Laurens and Geoffrey Hinton [2008]. *Visualizing data using t-SNE.* Journal of machine learning research 9.11 (2008) (cited on page 4).